



SEQUENCE LISTING

RECEIVED
AUG 23 2001
TECH CENTER 1600/2900

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: DLO-Center for Plant Breeding and
Reproduction Research
- (B) STREET: Droevendaalsesteeg 1
- (C) CITY: Wageningen
- (E) COUNTRY: The Netherlands
- (F) POSTAL CODE (ZIP): 6708 PB
- (G) TELEPHONE: +31 317 477001
- (H) TELEFAX: +31 317 418094
- (I) TELEX: -

(ii) TITLE OF INVENTION: A method for plant protection against insects
or nematodes

(iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/NL98/00352

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 888 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Actinia equina

(ix) FEATURE:

- (A) NAME/KEY: mat_peptide
- (B) LOCATION: 99..695

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..695

```
(ix) FEATURE:
      (A) NAME/KEY: sig_peptide
      (B) LOCATION:3..98
```

(B) LOCATION: 3..98

```
(ix) FEATURE:
      (A) NAME/KEY: 5'UTR
      (B) LOCATION:1..2
```

(B) LOCATION:1..2

```
(ix) FEATURE:
      (A) NAME/KEY: 3'UTR
      (B) LOCATION:696..888
```

(B) LOCATION: 696..888

(x) PUBLICATION INFORMATION:

(A) AUTHORS: Gruden, Kristina
Strukelj, Borut
Popovic, Tatjana
Lenarcic, Brigita
Bevec, Tadeja
Brzin, Joze
Kregar, Igor
Herzog-Velikonja, Jana
Stiekema, Willem J
Bosch, Dirk

(B) TITLE: The cysteine protease activity of Colorado potato beetle (*Leptinotarsa decemlineata*) guts, which is insensitive to potato protease inhibitors, is inhibited by thyroglobulin type-1 domain inhibitors

(C) JOURNAL: Insect Biochem. Mol. Biol

(D) VOLUME: 28

(F) PAGES: 549-560

(G) DATE: 1998

Bosch, Dirk

(B) TITLE: The cysteine protease activity of Colorado potato beetle (*Leptinotarsa decemlineata*) guts, which is insensitive to potato protease inhibitors, is inhibited by thyroglobulin type-1 domain inhibitors

(C) JOURNAL: Insect Biochem. Mol. Biol

(D) VOLUME: 28

(F) PAGES: 549-560

(G) DATE: 1998

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:															
ATG	GCT	CTT	AGC	CAA	AAC	CAA	GCC	AAG	TTT	TCC	AAA	GGA	TTC	GTC	47
Met	Ala	Leu	Ser	Gln	Asn	Gln	Ala	Lys	Phe	Ser	Lys	Gly	Phe	Val	
32		-30					-25					-20			
ATG	ATT	TGG	GTA	CTA	TTC	ATT	GCT	TGT	GCT	ATA	ACT	TCA	ACT	GAA	95
Met	Ile	Trp	Val	Leu	Phe	Ile	Ala	Cys	Ala	Ile	Thr	Ser	Thr	Glu	
	-15					-10					-5				
AGT	CTA	ACC	AAA	TGC	CAA	CAG	CTC	CAG	GCC	TCG	GCT	AAC	AGT	GGT	143
Ser	Leu	Thr	Lys	Cys	Gln	Gln	Leu	Gln	Ala	Ser	Ala	Asn	Ser	Gly	
1				5					10					15	
ATA	GGT	ACT	TAT	GTA	CCA	CAA	TGC	AAA	GAA	ACG	GGA	GAG	TTC	GAA	191
Ile	Gly	Thr	Tyr	Val	Pro	Gln	Cys	Lys	Glu	Thr	Gly	Glu	Phe	Glu	
			20					25					30		
AAA	CAA	TGC	TGG	GGA	TCG	ACT	GGT	TAC	TGT	TGG	TGT	GTG	GAT	GAA	239
Lys	Gln	Cys	Trp	Gly	Ser	Thr	Gly	Tyr	Cys	Trp	Cys	Val	Asp	Glu	
		35					40					45			

CT ATG GCT CTT AGC CAA AAC CAA GCC AAG TTT TCC AAA GGA TTC GTC 47
Met Ala Leu Ser Gln Asn Gln Ala Lys Phe Ser Lys Gly Phe Val
-32 -30 -25 -20

GTG ATG ATT TGG GTA CTA TTC ATT GCT TGT GCT ATA ACT TCA ACT GAA 95
Val Met Ile Trp Val Leu Phe Ile Ala Cys Ala Ile Thr Ser Thr Glu
-15 -10 -5

GCT AGT CTA ACC AAA TGC CAA CAG CTC CAG GCC TCG GCT AAC AGT GGT 143
Ala Ser Leu Thr Lys Cys Gln Gln Leu Gln Ala Ser Ala Asn Ser Gly
1 5 10 15

CTG ATA GGT ACT TAT GTA CCA CAA TGC AAA GAA ACG GGA GAG TTC GAA 191
Leu Ile Gly Thr Tyr Val Pro Gln Cys Lys Glu Thr Gly Glu Phe Glu
20 25 30

GAA AAA CAA TGC TGG GGA TCG ACT GGT TAC TGT TGG TGT GTG GAT GAA 239
Glu Lys Gln Cys Trp Gly Ser Thr Gly Tyr Cys Trp Cys Val Asp Glu
35 40 45

GAT GGA AAA GAG ATT CTA GGA ACC AAG ATC CGT GGA TCT CCG GAT TGC Asp Gly Lys Glu Ile Leu Gly Thr Lys Ile Arg Gly Ser Pro Asp Cys 50 55 60	287
AGC CGC AGA AAA GCC GCG TTA ACA CTT TGC CAG ATG ATG CAA GCC ATC Ser Arg Arg Lys Ala Ala Leu Thr Leu Cys Gln Met Met Gln Ala Ile 65 70 75	335
ATT GTT AAT GTC CCT GGT TGG TGT GGC CCT CCA TCG TGT AAA GCT GAC Ile Val Asn Val Pro Gly Trp Cys Gly Pro Pro Ser Cys Lys Ala Asp 80 85 90 95	383
GGC AGT TTT GAC GAG GTT CAG TGC TGC GCA AGT AAT GGA GAA TGC TAC Gly Ser Phe Asp Glu Val Gln Cys Cys Ala Ser Asn Gly Glu Cys Tyr 100 105 110	431
TGT GTG GAT AAG AAA GGA AAA GAA CTT GAA GGC ACA AGA CAA CAG GGA Cys Val Asp Lys Lys Gly Lys Glu Leu Glu Gly Thr Arg Gln Gln Gly 115 120 125	479
AGG CCA ACC TGC GAA AGA CAC CTA AGC GAA TGC GAG GAA GCT CGA ATC Arg Pro Thr Cys Glu Arg His Leu Ser Glu Cys Glu Glu Ala Arg Ile 130 135 140	527
AAG GCG CAT TCA AAC AGT CTT CGT GTT GAG ATG TTC GTG CCA GAG TGT Lys Ala His Ser Asn Ser Leu Arg Val Glu Met Phe Val Pro Glu Cys 145 150 155	575
TTA GAA GAT GGA TCA TAT AAC CCA GTA CAG TGC TGG CCT AGC ACA GGA Leu Glu Asp Gly Ser Tyr Asn Pro Val Gln Cys Trp Pro Ser Thr Gly 160 165 170 175	623
TAC TGT TGG TGC GTC GAT GAA GGA GGG GTA AAG GTA CCA GGT TCC GAT Tyr Cys Trp Cys Val Asp Glu Gly Gly Val Lys Val Pro Gly Ser Asp 180 185 190	671
GTC AGA TTT AAA CGC CCC ACA TGC TAAGAAAAAC ACAGTGAACA AAGTGGCTAG Val Arg Phe Lys Arg Pro Thr Cys 195	725
TTTCCAGATC GAAAATAACT ACAAAGGATT AATAAAATGT TAAAATAATT TCTCAATTTCG	785
GCTGTGATAT ATTTTTCCTCA AGATAATTTA ATCTGCATGT AGTTAACAGA AAACAATCTC	845
AACTAGAAAT AAAGACTACG GTAATAATGA CAAAAAAAAA AAA	888

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 231 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Leu Ser Gln Asn Gln Ala Lys Phe Ser Lys Gly Phe Val Val
-32 -30 -25 -20

Met Ile Trp Val Leu Phe Ile Ala Cys Ala Ile Thr Ser Thr Glu Ala
-15 -10 -5

Ser Leu Thr Lys Cys Gln Gln Leu Gln Ala Ser Ala Asn Ser Gly Leu
1 5 10 15

Ile Gly Thr Tyr Val Pro Gln Cys Lys Glu Thr Gly Glu Phe Glu Glu
20 25 30

Lys Gln Cys Trp Gly Ser Thr Gly Tyr Cys Trp Cys Val Asp Glu Asp
35 40 45

Gly Lys Glu Ile Leu Gly Thr Lys Ile Arg Gly Ser Pro Asp Cys Ser
50 55 60

Arg Arg Lys Ala Ala Leu Thr Leu Cys Gln Met Met Gln Ala Ile Ile
65 70 75 80

Val Asn Val Pro Gly Trp Cys Gly Pro Pro Ser Cys Lys Ala Asp Gly
85 90 95

Ser Phe Asp Glu Val Gln Cys Cys Ala Ser Asn Gly Glu Cys Tyr Cys
100 105 110

Val Asp Lys Lys Gly Lys Glu Leu Glu Gly Thr Arg Gln Gln Gly Arg
115 120 125

Pro Thr Cys Glu Arg His Leu Ser Glu Cys Glu Glu Ala Arg Ile Lys
130 135 140

Ala His Ser Asn Ser Leu Arg Val Glu Met Phe Val Pro Glu Cys Leu
145 150 155 160

Glu Asp Gly Ser Tyr Asn Pro Val Gln Cys Trp Pro Ser Thr Gly Tyr
165 170 175

Cys Trp Cys Val Asp Glu Gly Gly Val Lys Val Pro Gly Ser Asp Val
180 185 190

Arg Phe Lys Arg Pro Thr Cys
195

```

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 696 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Actinia equina

(vii) IMMEDIATE SOURCE:
    (B) CLONE: optimized gene for expression in plants

(ix) FEATURE:
    (A) NAME/KEY: CDS
    (B) LOCATION:1..693

(ix) FEATURE:
    (A) NAME/KEY: mat_peptide
    (B) LOCATION:97..693

(ix) FEATURE:
    (A) NAME/KEY: sig_peptide
    (B) LOCATION:1..693

```

ATG Met -32	GCT Ala	CTT Leu -30	AGC Ser	CAG Gln	AAC Asn	CAG Gln	GCC Ala -25	AAG Lys	TTT Phe	TCC Ser	AAG Lys	GGA Gly -20	TTC Phe	GTC Val	GTG Val	48
ATG Met	ATT Ile -15	TGG Trp	GTA Val	CTA Leu	TTC Phe	ATT Ile -10	GCT Ala	TGT Cys	GCT Ala	ATC Ile	ACT Thr -5	TCA Ser	ACT Thr	GAA Glu	GCT Ala	96
AGT Ser 1	CTA Leu	ACG Thr	AAA Lys	TGC Cys 5	CAA Gln	CAG Gln	CTG Leu	CAG Gln	GCC Ala 10	TCG Ser	GCT Ala	AAC Asn	AGT Ser	GGT Gly 15	CTG Leu	144
ATA Ile	GGT Gly	ACT Thr	TAT Tyr 20	GTA Val	CCA Pro	CAA Gln	TGC Cys	AAA Lys 25	GAA Glu	ACT Thr	GGA Gly	GAG Glu	TTT Phe 30	GAA Glu	GAA Glu	192
AAG Lys	CAA Gln 35	TGC Cys	TGG Trp	GGA Gly	TCG Ser	ACT Thr 40	GGT Gly	TAC Tyr	TGT Cys	TGG Trp	TGT Cys 45	GTG Val	GAT Asp	GAA Glu	GAT Asp	240

GGA AAA GAG ATT CTA GGT ACA AAG ATC CGT GGA TCT CCA GAC TGC AGT	288
Gly Lys Glu Ile Leu Gly Thr Lys Ile Arg Gly Ser Pro Asp Cys Ser	
50 55 60	
CGC AGA AAA GCT GCC TTA ACA CTT TGC CAG ATG ATG CAA GCC ATC ATT	336
Arg Arg Lys Ala Ala Leu Thr Leu Cys Gln Met Met Gln Ala Ile Ile	
65 70 75 80	
GTG AAT GTC CCT GGT TGG TGT GGA CCT CCA TCA TGT AAA GCT GAC GGC	384
Val Asn Val Pro Gly Trp Cys Gly Pro Pro Ser Cys Lys Ala Asp Gly	
85 90 95	
AGT TTT GAC GAG GTT CAG TGC TGC GCA AGT AAT GGA GAA TGC TAC TGT	432
Ser Phe Asp Glu Val Gln Cys Cys Ala Ser Asn Gly Glu Cys Tyr Cys	
100 105 110	
GTG GAT AAG AAA GGA AAA GAA CTT GAA GGC ACA AGA CAA CAG GGA AGG	480
Val Asp Lys Lys Gly Lys Glu Leu Glu Gly Thr Arg Gln Gln Gly Arg	
115 120 125	
CCA ACC TGC GAA AGA CAC CTA AGC GAA TGC GAG GAG GCT CGT ATC AAG	528
Pro Thr Cys Glu Arg His Leu Ser Glu Cys Glu Glu Ala Arg Ile Lys	
130 135 140	
GCA CAT TCA AAC AGT CTT CGT GTT GAG ATG TTC GTG CCA GAG TGT TTA	576
Ala His Ser Asn Ser Leu Arg Val Glu Met Phe Val Pro Glu Cys Leu	
145 150 155 160	
GAA GAT GGA TCT TAC AAC CCT GTA CAG TGC TGG CCT AGC ACA GGA TAC	624
Glu Asp Gly Ser Tyr Asn Pro Val Gln Cys Trp Pro Ser Thr Gly Tyr	
165 170 175	
TGT TGG TGC GTC GAT GAA GGA GGG GTA AAG GTT CCA GGT TCC GAC GTC	672
Cys Trp Cys Val Asp Glu Gly Gly Val Lys Val Pro Gly Ser Asp Val	
180 185 190	
AGA TTC AAA CGT CCC ACA TGC TAA	696
Arg Phe Lys Arg Pro Thr Cys	
195	

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 231 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Ala Leu Ser Gln Asn Gln Ala Lys Phe Ser Lys Gly Phe Val Val
-32 -30 -25 -20

Met Ile Trp Val Leu Phe Ile Ala Cys Ala Ile Thr Ser Thr Glu Ala
-15 -10 -5

Ser Leu Thr Lys Cys Gln Gln Leu Gln Ala Ser Ala Asn Ser Gly Leu
1 5 10 15

Ile Gly Thr Tyr Val Pro Gln Cys Lys Glu Thr Gly Glu Phe Glu Glu
20 25 30

Lys Gln Cys Trp Gly Ser Thr Gly Tyr Cys Trp Cys Val Asp Glu Asp
35 40 45

Gly Lys Glu Ile Leu Gly Thr Lys Ile Arg Gly Ser Pro Asp Cys Ser
50 55 60

Arg Arg Lys Ala Ala Leu Thr Leu Cys Gln Met Met Gln Ala Ile Ile
65 70 75 80

Val Asn Val Pro Gly Trp Cys Gly Pro Pro Ser Cys Lys Ala Asp Gly
85 90 95

Ser Phe Asp Glu Val Gln Cys Cys Ala Ser Asn Gly Glu Cys Tyr Cys
100 105 110

Val Asp Lys Lys Gly Lys Glu Leu Glu Gly Thr Arg Gln Gln Gly Arg
115 120 125

Pro Thr Cys Glu Arg His Leu Ser Glu Cys Glu Glu Ala Arg Ile Lys
130 135 140

Ala His Ser Asn Ser Leu Arg Val Glu Met Phe Val Pro Glu Cys Leu
145 150 155 160

Glu Asp Gly Ser Tyr Asn Pro Val Gln Cys Trp Pro Ser Thr Gly Tyr
165 170 175

Cys Trp Cys Val Asp Glu Gly Gly Val Lys Val Pro Gly Ser Asp Val
180 185 190

Arg Phe Lys Arg Pro Thr Cys
195